

## Beschreibung

## Identifizieren pharmazeutischer Targets

- 5 Die menschliche Erbsubstanz (Genom) umfasst schätzungsweise 20.000 bis 80.000 Gene, die den genetischen Code für etwa eine Million Eiweißstoffe (Proteine) beinhalten. In den spezialisierten Körperzellen werden jeweils nur Untermengen aller Gene tatsächlich abgelesen (exprimiert). Die Gesamtheit der
- 10 dadurch erzeugten Proteine wird als Proteom dieser Zelle bezeichnet. Das Wechselspiel der Proteine untereinander sowie mit der DNA stellt den wichtigsten Teil der Maschinerie dar, die der Entwicklung des menschlichen Körpers aus der befruchteten Eizelle sowie allen Körperfunktionen zugrunde liegt.
- 15 Aus der Sicht der Informatik stellt die Erbsubstanz damit einen prozeduralen Code für die Struktur und Funktion des menschlichen Körpers dar.

- Viele Krankheiten und Fehlfunktionen des Körpers gehen auf
- 20 Störungen des funktionellen Netzwerks aus Genom und Proteom zurück. So wirken einige Medikamente als Agonisten bzw. Antagonisten spezifischer Zielproteine, d. h. sie verstärken oder schwächen die Funktion eines Proteins mit dem Ziel, das aus Proteom und Genom gebildete regulatorische Netzwerk zurück in
- 25 einen normalen Funktionsmodus zu bringen. Diese Zielproteine (Targets) werden bislang nach heuristischen Prinzipien aus biochemischen Überlegungen abgeleitet. Oft ist dabei unklar, ob die Fehlfunktion eines Proteins tatsächlich die Krankheitsursache oder nur eines der Symptome einer versteckten
- 30 Fehlregulation an anderer Stelle des Netzwerks darstellt.

Aus [1] und [5] ist bekannt, Nervenzellen (Neuronen) und deren biologische Funktionalität, d.h. deren biologisches Verhalten, durch künstliche Neuronen nachzubilden.

35

Ferner ist aus [1] bekannt, zwei Arten von Synapsen bzw. Nervenzellen zu unterscheiden, eine erregende (exzitatorische)

## 2

Synapse 251 bzw. Nervenzelle und eine hemmende (inhibitorische) Synapse 252 bzw. Nervenzelle.

5 Hemmende Synapsen 252 senken zu übertragende bzw. weiterzu-  
leitende elektrische Potentiale, erregende Synapsen 251 erhöhen zu übertragende elektrische Potentiale.

Weitere Ausführungen zum Aufbau und zur Funktionalität einer Nervenzelle sowie zur Nervenleitung sind in [1] gegeben.

10

Ferner ist aus [1] und [5] eine künstliche Nervenzelle (künstliches Neuron), welche eine (biologische) Nervenzelle nachbildet bekannt.

15 Anschaulich gesehen ist ein solches künstliches Neuron eine mathematische Abbildung, welche entsprechend dem Übertragungsverhalten der biologischen Nervenzelle eine Eingangsgröße des künstlichen Neurons auf eine Ausgangsgröße des künstlichen Neurons abbildet.

20

In Anlehnung an das biologische Vorbild besteht ein künstliches Neuron aus drei Komponenten: dem Zellkörper, den Dendriten, welche Eingangssignale in das künstliche Neuron aufsummieren, dem Axon, welches das Ausgangssignal des künstlichen  
25 Neurons nach außen weiterleitet, sich verzweigt und mit den Dendriten nachfolgender künstlicher Neuronen über Synapsen in Kontakt tritt.

Eine Stärke einer Synapse bzw. die Art der Synapse wird meist  
30 durch einen numerischen Wert bzw. dessen Vorzeichen dargestellt. Dieser Wert wird als Verbindungsgewicht bezeichnet.

Entsprechend dem biologischen Vorbild lässt sich ein Übertragungsverhalten bzw. das Abbildungsverhalten eines künstlichen  
35 Neurons wie in [1] und [5] beschrieben abbilden.

Weitere Ausführungen zu künstlichen Neuronen und deren Funktionalität sind in [1] und [5] gegeben.

5      Ferner ist aus [2] und [5] bekannt, einzelne Neuronen miteinander zu verknüpfen. Eine solche Anordnung miteinander verknüpfter Neuronen wird als neuronales Netz bezeichnet. Grundlagen über neuronale Netze, beispielsweise verschiedene Arten von neuronalen Netzen, Trainingsverfahren für neuronale Netze, Bezüge zu biologischen Nervenzellen-Anordnungen, sind in  
10      [2] beschrieben.

Aus [3] und [5] ist eine Anwendung eines Mean-Field Modells bei der Beschreibung eines komplexen Systems bekannt. Bei einem Mean-Field Modell werden stochastische Wechselwirkungseinflüsse zwischen Komponenten eines Systems durch einen  
15      mittleren Wechselwirkungseinfluss genähert. Damit lassen sich analytisch nicht beschreibbare, stochastische Systeme auf beschreibbare, deterministische Systeme reduzieren.

20      Aus [4] und [5] ist die Anwendung des Mean-Field Modells bei der Beschreibung einer Neuronenstruktur bekannt.

Aufgabe der Erfindung ist es, das Identifizieren von Proteinen, die sich als Ziel medikamentöser Behandlung genetisch  
25      bedingter Krankheiten oder Störungen eignen, zu verbessern.

Diese Aufgabe wird durch die Erfindungen gemäß den unabhängigen Ansprüchen gelöst. Vorteilhafte Weiterbildungen der Erfindungen sind in den Unteransprüchen gekennzeichnet.

30

Erfindungsgemäß wird eine Mehrzahl von Genexpressionsmustern gleichartiger Zellen oder eines Gewebes bestimmt, wobei jeweils die Expressionsrate der Gene der Zelle bestimmt wird. Die Mehrzahl der Genexpressionsmuster wird derart bestimmt,  
35      dass der zeitliche Verlauf der Genexpressionsmuster der Zelle zumindest teilweise rekonstruiert werden kann. Ein dynamisches Modell des regulatorischen Netzwerks aus Genom und Pro-

teom der Zelle wird gebildet, indem ein äquivalentes Neuronales Netz in der folgenden Weise gebildet wird:

i) Ein Gen des Genoms sowie das zugehörige Protein werden durch ein Neuron des äquivalenten Neuronalen Netzes repräsentiert.

ii) Die Expressionsrate eines Gens wird durch eine nicht-negative Aktivität des äquivalenten Neurons repräsentiert.

iii) Die regulatorische Wirkung eines Proteins auf ein Gen wird durch eine synaptische Verbindung von dem zum Protein äquivalenten Neuron zu dem zum Gen äquivalenten Neuron repräsentiert.

iv) Die Art einer regulatorischen Wirkung (verstärkend oder inhibitorisch) wird im Neuronalen Netz durch das Vorzeichen und die Stärke eines zugehörigen synaptischen Gewichts repräsentiert.

v) In einer Weiterentwicklung ist es auch möglich, eine posttranslationale Modifikation eines ersten Proteins durch ein zweites Protein durch eine synaptische Verbindungen mit multiplikativer Wirkung vom zweiten Neuron zum ersten Neuron zu repräsentieren.

vi) In einer weiteren Weiterentwicklung kann ein externer Einfluss auf das regulatorische Netzwerk durch einen Inputknoten des äquivalenten Neuronalen Netzes repräsentiert werden.

Das äquivalente Neuronale Netz wird mit den bestimmten Genexpressionsmustern verglichen und an diese angepasst. Aus dem angepassten Neuronalen Netz wird das regulatorische Netzwerk der untersuchten Zelle erschlossen.

Es wird also eine Äquivalenzbeziehung zwischen dem funktionellen Netzwerk des Genoms und Proteoms einerseits und dem Neuronalen Netz des menschlichen Gehirns andererseits, die beide stark vernetzte rückgekoppelte Systeme darstellen, hergestellt. Durch diese Abbildung gelingt die Modellierung des funktionellen Netzwerks aus Proteinen und Genen.

Das Verfahren erlaubt die Identifizierung von Targetproteinen auf systematischer Basis. Zwischen genetischen und Neuronalen Netzen kann die beschriebene Äquivalenzbeziehung hergestellt werden. Das dynamische Wechselspiel von Genen und regulatorischen Proteinen wird damit durch ein dynamisches Neuronales Netz modelliert. Das Verfahren nutzt die in der Zeitabfolge der Genexpressionsmuster enthaltene Information für die Identifizierung kausaler regulatorischer Zusammenhänge.

10 In der Regel führen genetisch bedingte Krankheiten zu komplexen Fehlfunktionen, die aber oft auf nur ein oder wenige fehlerhaft arbeitende Gene oder Proteine zurückgehen. Bislang sind diese Schlüsselgene bis auf Einzelfälle nicht bekannt. Stattdessen werden im herkömmlichen Target Finding durch heuristisches Vorgehen Zielproteine (Targets) gesucht, deren medikamentöse Regulierung das gesunde Genexpressionsmuster bestmöglich wiederherstellen soll.

20 Neuartige Schätzungen sprechen von etwa 10.000 verschiedenen Proteinen als mögliche Targets im menschlichen Genom, deren Durchforstung durch einen heuristischen Ansatz allein nicht zu bewerkstelligen ist.

25 Der hier beschriebene Modellansatz stellt ein leistungsfähiges Verfahren zum systematischen Target Finding dar, also zur Identifikation eines oder weniger Schlüsselgene oder -proteine, die am Anfang der Regulationskaskade stehen und die beispielsweise eine Organentwicklung einleiten, die Regenerationsfähigkeit von Gewebe regulieren oder aber auch für 30 meist komplexe Veränderungen von Genexpressionsmustern im Krankheitsfall verantwortlich zeichnen.

Das beschriebene Verfahren erlaubt ein computerunterstütztes Target Finding, das in der Lage ist, die große Datenmenge und 35 die vielfältigen und komplexen Zusammenhänge zu analysieren.

Es lassen sich damit folgende Anwendungsgebiete spezifizieren:

- Modellbasierte Unterstützung von Forschungsaktivitäten zur Entschlüsselung der menschlichen Morphogenese und damit der generellen Prinzipien von genetisch gesteuerten Wachstums-, Regenerations- und Abbauprozessen.

- Unterstützung der Identifikation von Zielproteinen, die für Wachstumsstörungen wie z. B. ungehemmtes Tumorwachstum ursächlich verantwortlich sind und nicht nur ein Symptom darstellen. Daraus können neuartige Verfahren für eine hochsensitive Tumorfrühdiagnostik aber auch Behandlungsmethoden wie der selektiv induzierte Zelltod (Apoptose) in Tumorzellen abgeleitet werden.

- Unterstützung der Identifikation von regulatorischen Proteinen, die in Wachstums- und Regenerationsprozesse eingreifen. Damit wäre eine wichtige Hürde auf dem Wege der induzierten Regeneration von Organen genommen.

Im Folgenden wird die Erfindung anhand von Ausführungsbeispielen näher erläutert, die in den Figuren schematisch dargestellt sind. Gleiche Bezugsziffern in den einzelnen Figuren bezeichnen dabei gleiche Elemente. Im Einzelnen zeigt:

Fig. 1 schematisch die regulatorischen Vorgänge, die das Expressionsmuster einer Zelle bestimmen;

Fig. 2 eine schematische Darstellung der Vernetzung von Neuronen;

Fig. 3 die Potenziale innerhalb eines Dendriten bzw. eines Neurons als Funktion der Zeit; und

Fig. 4 eine schematische Darstellung einer modulatorischen Synapse.

Fig. 1 zeigt die wichtigsten Wechselwirkungen zwischen Genen und Proteinen eines DNA-Abschnitts auf. Die Wechselwirkungen werden als Basis für die Beschreibung des genomischen regulatorischen Netzwerks herangezogen.

5

Im oberen Teil der Fig. 1 ist schematisch angedeutet, wie ein von außen auf die Zelle einwirkendes externes Signal - etwa im Rahmen der interzellulären Kommunikation -, das beispielsweise von einem Transmembran-Rezeptorprotein (z. B. von einem Kalziumkanal) aufgenommen und in geeigneter Weise in das Innere der Zelle übertragen wird, die Produktion der Gene A, B, C und D des DNA-Abschnitts auslöst.

10

15

Es besteht daher prinzipiell auch die Möglichkeit, die Expressionsrate einzelner Gene einer Zelle über die erwähnten Wege von außerhalb der Zellen zu beeinflussen.

20

Als ein Gen wird ein nicht notwendigerweise zusammenhängender Abschnitt der DNA bezeichnet, der den genetischen Code für ein Protein oder auch für eine Gruppe von Proteinen enthält. Im allgemeinen weist die DNA sogenannte Exons und Introns auf. Exons stellen Teile der DNA dar, die tatsächlich ein Protein codieren. Introns stellen Teile der DNA dar, die nicht unmittelbar ein Protein codieren. In erster Näherung sind sie ohne Funktion. Exons und Introns wechseln einander in der DNA ab. Bezeichnet man als ein Gen die Menge der Exons, die zusammen ein bestimmtes Protein codieren, so ist ein solches Gen - wie oben erwähnt - in der Regel nicht zusammenhängend.

25

30

Der Produktionsvorgang eines Proteins aus einem Gen, zum Beispiel Protein A ausgehend von Gen A in Fig. 1, wird als Expression dieses Gens bezeichnet. Die Übersetzung des DNA-Codes des Gens in die Kette der Aminosäuren des Proteins wird als Translation bezeichnet. Die Rate, mit der das Protein A in einem gegebenen Kontext produziert wird, wird seine Expressionsrate genannt.

35

Nicht alle Gene werden in einer Zelle exprimiert. Vielmehr unterscheiden sich verschiedene Zelltypen durch ihr Genexpressionsmuster. Dies gilt oftmals auch für den Unterschied zwischen kranken und gesunden Zellen.

Das Expressionsmuster einer Zelle wird durch die in Fig. 1 schematisch dargestellten regulatorischen Vorgänge bestimmt. Die regulatorischen Vorgänge werden im Wesentlichen von einigen wichtigen Wechselwirkungen zwischen Proteinen und Genen sowie zwischen den Proteinen untereinander bestimmt.

So kann die Expressionsrate eines Gens A durch das Vorhandensein eines anderen Proteins B reguliert, d. h. erhöht, erniedrigt oder zum Erliegen gebracht werden. In diesem Beispiel wirkt das Protein B regulatorisch auf das Gen A bzw. das Protein A. Zu regulatorischen Proteinen können beispielsweise die Proteinbausteine von Aktivatorkomplexen gerechnet werden. Regulatorische Proteine können sich gleichzeitig auf viele Zielgene auswirken.

Eine zweite Art der Wechselwirkung besteht in der posttranslationalen Modifikation von Proteinen, d. h. der Modifikation von Proteinen nach der Translation. In der Regel erfolgt die posttranslationale Modifikation eines Proteins im unmittelbaren Anschluss an die Translation, d. h. bevor das Protein in der Zelle wirkt. So werden zum Beispiel viele Proteine von speziellen Enzymen phosphoryliert oder glykolyliert, d. h. das Zielprotein wird durch Anhängen bzw. Abspalten chemischer Gruppen in seinen funktionellen Zustand gebracht oder in einen Zustand versetzt, in dem es nicht mehr wirksam ist. Posttranslationale Modifikation kann also ein Protein gegebenenfalls temporär funktionell einschalten oder ausschalten.

In Fig. 1 ist das Protein A ein sogenanntes Effektorprotein, d. h. es wirkt innerhalb der Zelle auf andere Substanzen und nicht unmittelbar auf das Genom oder Proteom. In Fig. 1 modi-



fiziert das Protein C im Wege der posttranslationalen Modifikation die Funktion des Effektorproteins A.

Protein B ist ein regulatorisches Protein, da es die Expressionsrate des Proteins A bestimmt, indem es mit demjenigen DNA-Abschnitt wechselwirkt, der das Gen A enthält. Das Protein D modifiziert somit die Funktion eines regulatorischen Proteins (Protein B) im Wege der posttranslationalen Modifikation.

10

Datenbasis

Die Nukleinsäuresequenz der menschlichen DNA ist weitestgehend bekannt. Auch die durch die DNA codierten Gene sind in zunehmendem Maße identifiziert. Nicht ganz so vollständig ist das Wissen über das Proteom, einschließlich der eventuell durch Wechselwirkung zwischen den Proteinen posttranslational modifizierten Proteine. Allerdings erlauben neuere Sequenzierungs- und Hochdurchsatz-Screeningverfahren eine rasche Identifizierung weiterer Gene und Proteine.

Ein weiterer wichtiger Schritt zur Aufklärung der Expressionsmuster einer Zelle hat sich mit der Entwicklung von Hochdurchsatz-Hybridisierungstechniken vollzogen. Bei diesen Verfahren wird auf einem sogenannten Microarray die Expressionsrate vieler 100 verschiedener Gene gleichzeitig getestet. Mit Hilfe dieser Verfahren ist es möglich, das Genexpressionsmuster einer Zelle zu bestimmen.

Dazu werden in der Regel die in der Zelle synthetisierten mRNA (messenger RNA) bestimmt. Die mRNA ist ein Zwischenprodukt bei der Translation des Gens zum Protein. Die mRNA ist somit eine Vorstufe bei der Bildung des Proteins und weist auf die Bildung des zugehörigen Proteins hin. Die zu untersuchende Zelle wird zunächst isoliert. Anschließend wird sie aufgeschlossen. Durch geeignete Aufreinigungsschritte wird die mRNA aus der Zelle isoliert. Danach wird die mRNA mittels

der reversen Transkriptase in cDNA (complementary DNA) übersetzt. Diese wird mit i. d. R. linearer PCR (polymerase chain reaction) amplifiziert. Die so gewonnene cDNA wird mit Hilfe von geeigneten Microarrays, z. B. DNA-Chips, qualitativ bzw. quantitativ analysiert. Mit modernen Microarrays können die Expressionsraten von 5.000 und mehr Genen gleichzeitig gemessen werden.

Aufgrund dieser verbesserten Techniken steht mittlerweile ein umfangreiches Wissen über das menschliche Genom und Proteom sowie über die Wechselwirkungen zwischen Proteinen und Genen bzw. Proteinen untereinander zur Verfügung.

Von besonderer Bedeutung sind Datensätze, in denen die zeitliche Abfolge von Genexpressionsmustern in einem Gewebe festgehalten ist. Als Beispiel sind sogenannte longitudinale Hybridisierungsstudien, d. h. zeitliche Abfolgen von Genexpressionsmustern während der Organdifferenzierung im Rahmen der Embryonalentwicklung, zu nennen. Zeitaufgelöste Genexpressionsdaten existieren auch für den Zellteilungs-Zyklus von Einzellern und sind auch für komplexere Gewebe möglich.

Neuronale Modellierung von Genom und Proteom (vgl. [5])

Im Folgenden wird allgemein das Modellierungsprinzip geschildert. Grundzüge sind aus [5] bekannt.

Das Grundprinzip besteht in der Herstellung einer Äquivalenzbeziehung zwischen dem funktionellen Netzwerk des Genoms und Proteoms einerseits und dem Neuronalen Netz des menschlichen Gehirns andererseits, die beide stark vernetzte rückgekoppelte Systeme darstellen.

Im Folgenden wird das Neuronale Netz des menschlichen Gehirns in einigen Grundzügen anhand von Fig. 2 geschildert.

Im menschlichen Gehirn befinden sich etwa hundert Milliarden Nervenzellen 20 (Neuronen), die jede mit Zehntausenden anderen Nervenzellen 20 Informationen austauscht. Die Informationen gelangen von einem Neuron 20 über das zu jedem Neuron gehörende Axon 22 zu einem anderen Neuron 20. Jedes Neuron hat genau ein Axon, um Informationen an andere Neuronen zu senden. In seinem weiteren Verlauf verzweigt sich ein Axon typischerweise etwa eintausend Mal, so dass ein Neuron 20 über sein Axon 22 Informationen an etwa eintausend andere Neuronen 20 senden kann.

Zum Empfangen von Informationen verfügen Neuronen 20 über Dendriten 24. Das die Information tragende Axon 22 ist über eine Synapse 26 mit dem Dendriten 24 verbunden. Über diese Synapse 26 gelangt die Information vom Axon 22 in den Dendrit 24 und damit vom aussendenden Neuron 20 zum nachgeschalteten Neuron. An einem einzigen Dendriten 24 können Tausende bis Hunderttausende Axons 22 bzw. Synapsen 26 angreifen, so dass ein nachgeschaltetes Neuron 20 Signale von vielen 1000 vorgeschalteten Neuronen empfangen kann.

Im Folgenden wird auf Fig. 3 Bezug genommen, die die Potenziale innerhalb eines Dendriten bzw. eines Neurons als Funktion der Zeit darstellt. Der Informationsaustausch zwischen den Neuronen 20 geschieht in Form von Aktionspotenzialen (Spikes) 30, die jedes Neuron 20 über sein Axon 22 aussendet. Die Spikes evozieren in den nachgeschalteten Neuronen 20 erneut Signale, die sogenannten postsynaptischen Potenziale (PSPs) 32. Die Größe dieser PSPs hängt von der Übertragungsstärke oder dem synaptischen Gewicht  $w$  der jeweiligen Synapse ab.

Im Folgenden wird auf Fig. 4 Bezug genommen. Fig. 4 zeigt einen Dendriten 24, an den eine erste Synapse 26 koppelt. An diese erste Synapse 26 ist eine zweite Synapse 36 gekoppelt. Diese zweite Synapse 36 wird als modulatorische Synapse bezeichnet. Bezeichnen wir, in Anlehnung an Fig. 3, mit PSP das postsynaptische Potenzial 32, dass sich im Dendriten 24 auf-

12

grund der Einwirkung der ersten Synapse 26 in Abwesenheit der modulatorischen Synapse 36 ausbilden würde, so kann dies dargestellt werden durch

$$PSP = w \cdot \varepsilon(t),$$

wobei  $w$ , wie oben definiert, das synaptische Gewicht der ersten Synapse 26 wiedergibt, und  $\varepsilon(t)$  den zeitlichen Verlauf des postsynaptischen Potenzials 32 in einer geeigneten Normierung wiedergibt.

Greift an die erste Synapse 26 zusätzlich die modulatorische Synapse 36 an, so ergibt sich im Dendriten 24 ein modifiziertes postsynaptisches Potenzial  $PSP'$ , das durch einen multiplikativen Term  $act$  ausgedrückt werden kann:

$$PSP' = act \cdot w \cdot \varepsilon(t) = act \cdot PSP.$$

Dabei bezeichnet  $act$  die Aktivität der modulatorischen Synapse 36.

Beispielsweise haben im Zentralnervensystem, also im Gehirn und Rückenmark, dopaminerge Synapsen modulatorischen Charakter.

Die neuronale Aktivität jedes Neurons, d. h. die Anzahl der pro Zeiteinheit ausgesendeten Spikes, ergibt sich - vereinfacht gesprochen - durch eine nichtlineare und zeitlich nichtlokale Funktion aller einlaufenden postsynaptischen Potentiale. Überschreitet diese Funktion einen bestimmten Schwellenwert, so wird ein Spike 30 ausgelöst und über das Axon 22 übertragen.

Somit stellt das biologische Neuronale Netz des Gehirns ein komplexes, nichtlineares System dar, das obendrein eine hohe Vernetzungsdichte aufweist. Für die modellhafte Beschreibung dieses Systems hat die Neuroinformatik in den vergangenen

Jahren leistungsfähige Theorien und Algorithmen entwickelt (z.B. Kompartimentmodelle, Spike-Reponse-Modelle, Mean-Field-Modelle, Multimodulare Neurokognitive Modelle, Bayes-Belief-Netze).

5

Diese Theorien bzw. Gleichungen entsprechen in ihrer Struktur den weiter oben für die Reaktionskinetik abgeleiteten Gleichungen. Daher kann das regulatorische Netzwerk aus Genom und Proteom einer Zelle wie folgt auf ein äquivalentes Neuronales

10 Netz abgebildet werden:

15

- Ein Gen A des Genoms (hier verstanden als diejenige Kombination von Exons, die eindeutig ein Protein codiert) sowie das zugehörige Protein A werden mit einem Neuron A des äquivalenten Neuronalen Netzes identifiziert. Da in Genexpressionsmustern nur die mRNA bzw. cDNA quantitativ analysiert wird, kann auf der Ebene der Genexpressionsmuster allein ohnehin nicht zwischen Genen und Proteinen unterschieden werden.

20

- Die Expressionsrate eines Gens A wird als nichtnegative Aktivität, z. B. die Spikerate des Neurons A aufgefasst.

25

- Wirkt ein Protein B regulatorisch auf ein nachgeschaltetes Gen A, so enthält das äquivalente Neuronale Netz eine synaptische Verbindung von Neuron B zu dem äquivalenten nachgeschalteten Neuron A.

30

- Die Art der regulatorischen Wirkung (verstärkend oder inhibitorisch) wird im Neuronalen Netz durch das Vorzeichen und die Stärke des zugehörigen synaptischen Gewichts angegeben.

35

- Eine posttranslationale Modifikation eines Proteins durch ein anderes Protein, in Fig. 1 z. B. die Modifikation von Protein B durch Protein D, entspricht der Wirkung einer modulatorischen Synapse im Zentralnervensystem. Modulatorische Synapsen werden in künstlichen Neuronalen Netzen durch synap-

tische Verbindungen mit multiplikativer Wirkung auf andere Synapsen beschrieben. Die äquivalente Widerspiegelung einer posttranslationalen Modifikation des Proteins B durch das Protein D ist also eine synaptische Verbindungen mit multiplikativer Wirkung von Neuron D zu Neuron B.

- Externe Signale werden mit Inputknoten des äquivalenten Neuronalen Netzes identifiziert.

10 Zwischen genetischen und Neuronalen Netzen kann die beschriebene Äquivalenzbeziehung hergestellt werden. Das dynamische Wechselspiel von Genen und regulatorischen Proteinen wird damit durch ein dynamisches Neuronales Netz modelliert.

15 Zu den geeigneten neuronalen Algorithmen zählen Netze spikender Neuronen aber auch Mean-Field Modelle, die den expliziten Zeitablauf der Signalübertragung zwischen den Neuronen durch die explizite Beschreibung der postsynaptischen Potenziale berücksichtigen. Sie erlauben die Modellierung der zeitlichen  
20 Entwicklung der neuronalen Aktivitäten im Netz als Ergebnis externer Stimulation oder intrinsischer Aktivität.

Der zeitliche Verlauf der Konzentrationen, der sich durch die Reaktionskinetik zwischen den beteiligten Molekülen (z. B.  
25 zwischen regulatorischem Protein und DNA-Promotor) ergibt, wird also durch den Zeitverlauf der Aktivitäten der Neuronen ersetzt, so dass das resultierende Netzmodell zur Simulation der zeitlichen Entwicklung von Genexpressionsmustern herangezogen werden kann.

30 Für ein solches Neuronales Netz können die neuronalen Aktivitäten über die Zeit erschlossen werden. Da die neuronalen Aktivitäten den Genexpressionsmustern entsprechen, können beide miteinander verglichen werden. Das Neuronale Netz entspricht  
35 einem simulierten Genexpressionsmuster.

Ziel der Modellierung ist es, das dem Expressionsverlauf zugrundeliegende regulatorische Netzwerk zu ermitteln, also die Frage zu beantworten: "Welche neuronale Vernetzungsstruktur mit welchen Gewichten und Reaktionskonstanten ist konsistent mit dem beobachteten Genexpressionsverlauf?"

Zur Beantwortung dieser Frage wird das Netz mit einem an Strukturlernen angelehnten Verfahren trainiert: Es wird versucht, das beobachtete Verhalten mit so wenigen regulatorischen Verbindungen wie möglich aber auch so gut wie möglich zu erklären, also das einfachste mit den Daten konsistente Modell zu finden.

Ein bevorzugtes Optimierungsverfahren minimiert die totale Abweichung zwischen gemessenen und simulierten Genexpressionsmustern unter Verwendung eines "sparse priors", also einer Zusatzbedingung, die die Koexistenz vieler Verbindungen mit kleinen Gewichten zugunsten weniger regulatorischer Verbindungen bestraft. Eine Möglichkeit, einen solchen "sparse prior" zu realisieren, ist der Fachwelt bekannt .

Kreuzvalidierung und statistische Optimierung erlauben die Abschätzung der Eindeutigkeit der Lösung sowie ihrer Vorhersagekraft (Generalisierungsfähigkeit).

Dem fertig trainierten Netz können anhand der Verbindungsstruktur der Neuronen kausale Zusammenhänge zwischen Genen aber auch die Rolle verschiedener Gene entnommen werden. So weist ein asymmetrisches Gewicht nur von Gen B zu Gen A darauf hin, dass Gen B Gen A reguliert. Gleichzeitig können im Modell verschiedene Gene oder regulatorische Verbindungen künstlich ab- oder angeschaltet und die Auswirkungen auf die Genexpressionsmuster mit dem Ziel quantifiziert werden, die Ursache(n) krankheitsbedingter Veränderungen der Genexpression zu identifizieren (sogenannte inverse Modellierung).

Im Rahmen dieses Dokuments sind folgende Veröffentlichungen zitiert:

- 5 [1] Zell, A., "Simulation Neuronaler Netze", S. 35 bis 51,  
S. 55 bis 86, Addison-Wesley Longman Verlag GmbH, 1994,  
3. Unveränderter Nachdruck, R. Oldenbourg Verlag, ISBN  
3-486-24350-0, 2000
- 10 [2] Zell, A., "Simulation Neuronaler Netze", S. 87 bis 114,  
Addison-Wesley Longman Verlag GmbH, 1994, 3. Unverän-  
derter Nachdruck, R. Oldenbourg Verlag, ISBN 3-486-  
24350-0, 2000
- 15 [3] J. J. Binney, N. J. Dowrick, A. J. Fisher, M. E. J. New-  
man, "The Theory of Critical Phenomena", Kap. 6: Mean-  
Field Theory, Clarendon Press Oxford 1992
- 20 [4] C. Koch, I. Segev (Hrsg), "Methods of Neural Modeling:  
From Ions to Networks", Kap 13: D. Hansel and H. Sompo-  
linsky: "Modeling Feature Selectivity in Local Circu-  
its", MIT Press, Cambridge, 1998
- [5] Patentanmeldung PCT/DE02/03381



## Patentansprüche

1. Verfahren zum Identifizieren pharmazeutischer Targets mit folgenden Schritten:

- 5 a) eine Mehrzahl von Genexpressionsmustern gleichartiger Zellen wird bestimmt, wobei jeweils die Expressionsrate der Gene der Zelle bestimmt wird;
- b) die Mehrzahl der Genexpressionsmuster wird derart bestimmt, dass der zeitliche Verlauf der Genexpressionsmuster  
10 der Zelle zumindest teilweise rekonstruiert werden kann;
- c) ein dynamisches Modell des regulatorischen Netzwerks aus Genom und Proteom der Zelle wird gebildet, indem ein äquivalentes Neuronales Netz in der folgenden Weise gebildet wird:
- 15 i) ein Gen des Genoms sowie das zugehörige Protein werden durch ein Neuron (20) des äquivalenten Neuronalen Netzes repräsentiert;
- ii) die Expressionsrate eines Gens wird durch eine nichtnegative Aktivität des äquivalenten Neurons repräsentiert;  
20 tiert;
- iii) die regulatorische Wirkung eines Proteins auf ein Gen wird durch eine synaptische Verbindung (26) von dem zum Protein äquivalenten Neuron zu dem zum Gen äquivalenten Neuron repräsentiert;
- 25 iv) die Art einer regulatorischen Wirkung (verstärkend oder inhibitorisch) wird im Neuronalen Netz durch das Vorzeichen und die Stärke eines zugehörigen synaptischen Gewichts (w) repräsentiert;
- d) das äquivalente Neuronale Netz wird mit den bestimmten  
30 Genexpressionsmustern verglichen und an diese angepasst;
- e) aus dem angepassten Neuronalen Netz wird das regulatorische Netzwerk der untersuchten Zelle erschlossen.

2. Verfahren nach dem vorhergehenden Anspruch,  
35 dadurch gekennzeichnet,  
dass eine posttranslationale Modifikation eines ersten Proteins durch ein zweites Protein durch eine synaptische

Verbindungen (36) mit multiplikativer Wirkung vom zweiten Neuron zum ersten Neuron repräsentiert wird.

3. Verfahren nach einem der vorhergehenden Ansprüche,  
dadurch gekennzeichnet,  
dass ein externer Einfluss auf das regulatorische Netzwerk durch einen Inputknoten des äquivalenten Neuronalen Netzes repräsentiert wird.

4. Verfahren nach einem der vorhergehenden Ansprüche,  
dadurch gekennzeichnet,  
dass bei der Anpassung des äquivalenten Neuronalen Netzes an die bestimmten Genexpressionsmuster solche Netze bevorzugt werden, die einen geringeren Grad an Vernetzung aufweisen.

5. Anordnung zum Identifizieren pharmazeutischer Targets mit Mitteln zum Bestimmen einer Mehrzahl von Genexpressionsmustern einer Zelle, wobei jeweils die Expressionsrate der Gene der Zelle bestimmt wird, wobei die Mittel derart ausgebildet sind, dass der zeitliche Verlauf der Genexpressionsmuster der Zelle zumindest teilweise rekonstruiert werden kann;

mit Mitteln zum Bilden eines dynamischen Modells des regulatorischen Netzwerks aus Genom und Proteom der Zelle, indem ein äquivalentes Neuronales Netz in der folgenden Weise gebildet wird:

i) Ein Gen des Genoms sowie das zugehörige Protein werden durch ein Neuron (20) des äquivalenten Neuronalen Netzes repräsentiert;

ii) die Expressionsrate eines Gens wird durch eine nichtnegative Aktivität des äquivalenten Neurons (20) repräsentiert;

iii) die regulatorische Wirkung eines Proteins auf ein Gen wird durch eine synaptische Verbindung (26) von dem zum Protein äquivalenten Neuron zu dem zum Gen äquivalenten Neuron repräsentiert;

iv) die Art einer regulatorischen Wirkung (verstärkend oder inhibitorisch) wird im Neuronalen Netz durch das Vorzeichen und die Stärke eines zugehörigen synaptischen Gewichts (w) repräsentiert;

5 mit Mitteln zum Vergleichen des äquivalenten Neuronalen Netzes mit den bestimmten Genexpressionsmustern;

mit Mitteln zum Anpassen des äquivalenten Neuronalen Netzes an die bestimmten Genexpressionsmuster; und

10 mit Mitteln zum Erschließen des regulatorischen Netzwerks der untersuchten Zelle aus dem angepassten Neuronalen Netz.

FIG 1

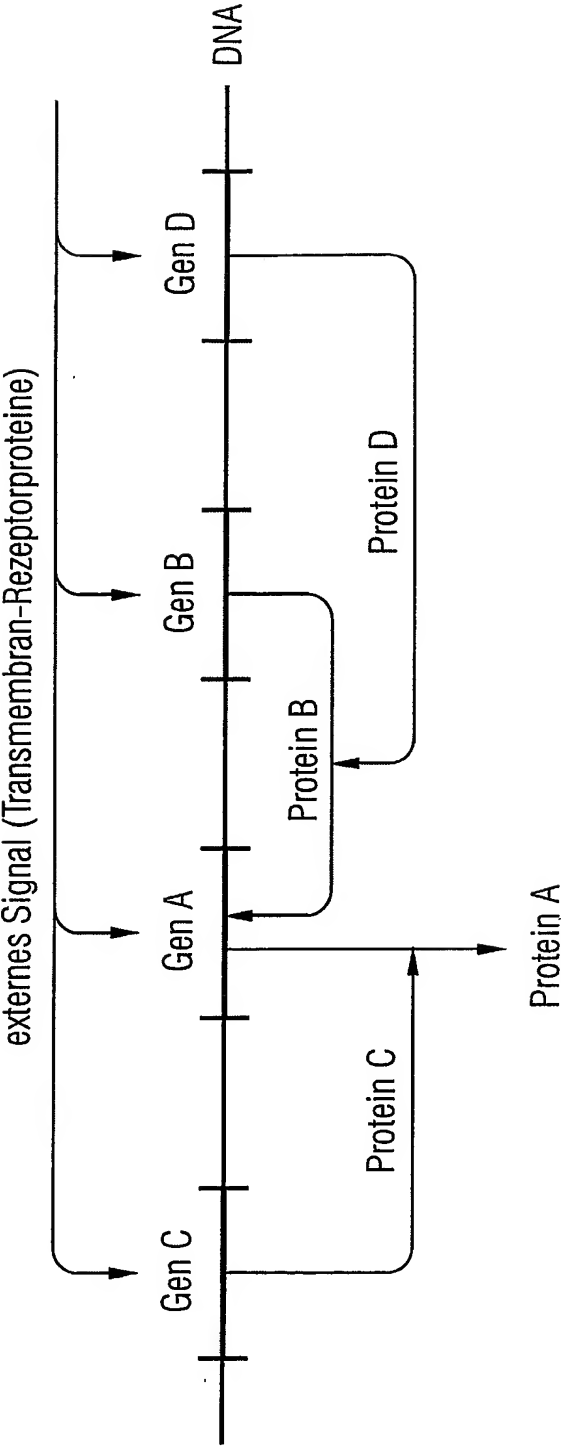


FIG 2

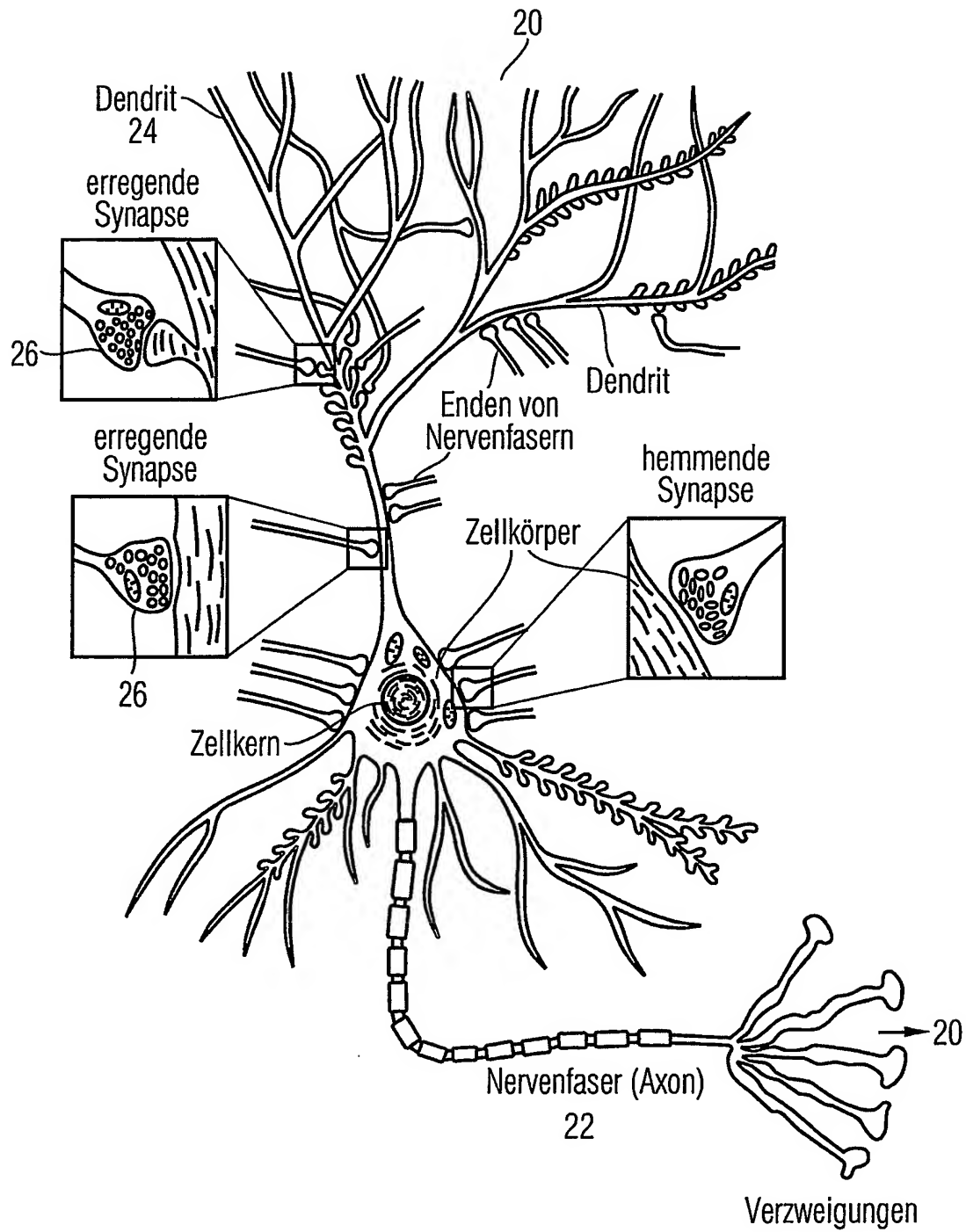


FIG 3

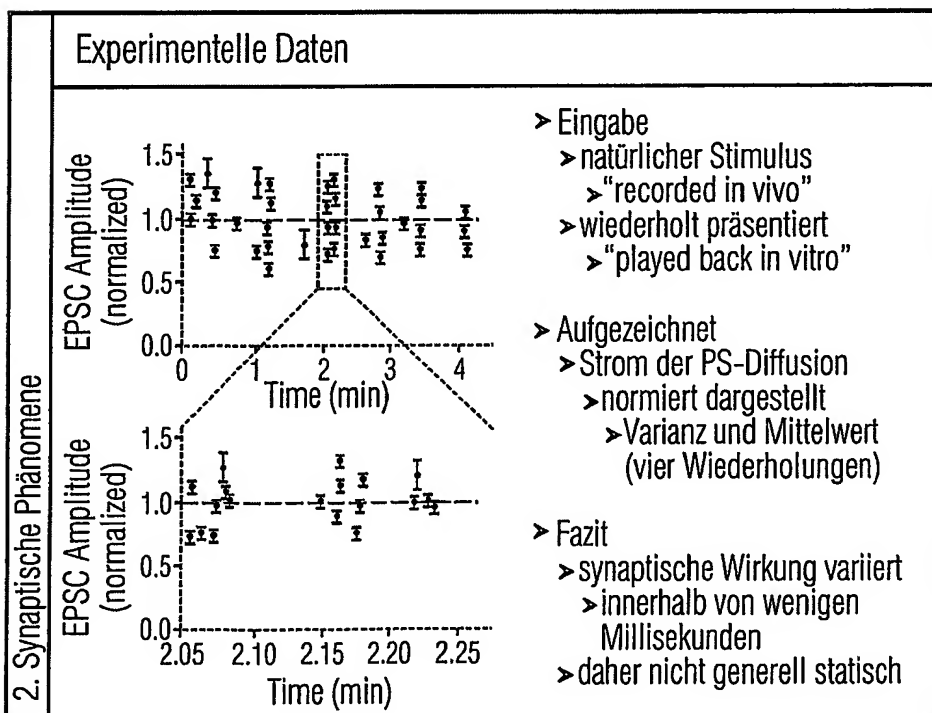
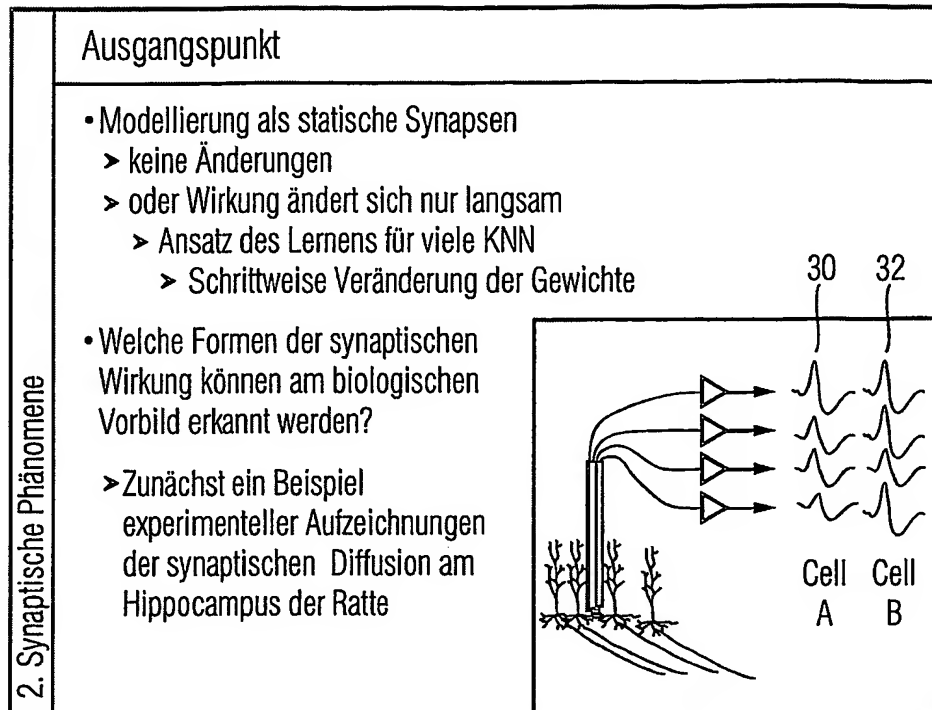
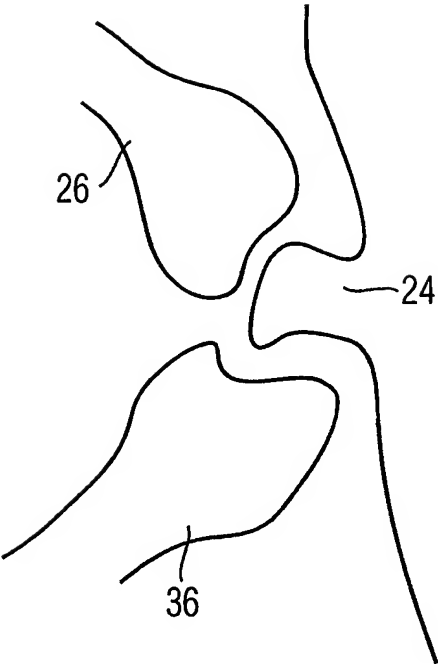


FIG 4



# INTERNATIONAL SEARCH REPORT

International Application No

PCT/EP2004/051835

**A. CLASSIFICATION OF SUBJECT MATTER**  
IPC 7 G01N33/50 G06F19/00

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)  
IPC 7 G01N G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2002/077756 A1 (DIAMOND CORNELIUS ET AL) 20 June 2002 (2002-06-20) paragraphs '0050!', '0051!', '0163!', '0217!; claims 10,21-24	1-5
X	WO 02/37102 A (CHILDRENS MEDICAL CENTER) 10 May 2002 (2002-05-10) paragraphs '0007!', '0014!', '0015!', '0077!; claims 11,38-42	1-5
X	WO 02/10453 A (PORTER MARK W ; CASTLE ARTHUR L (US); GENE LOGIC INC (US); JOHNSON KOR) 7 February 2002 (2002-02-07) page 34, paragraph 1-3 page 43, line 28; claims 51-54,29-43 abstract	1-5
	----- -/--	

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

\* Special categories of cited documents :

- \*A\* document defining the general state of the art which is not considered to be of particular relevance
- \*E\* earlier document but published on or after the International filing date
- \*L\* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- \*O\* document referring to an oral disclosure, use, exhibition or other means
- \*P\* document published prior to the international filing date but later than the priority date claimed

- \*T\* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- \*X\* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- \*Y\* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- \*8\* document member of the same patent family

Date of the actual completion of the international search

12 January 2005

Date of mailing of the international search report

20/01/2005

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
Fax: (+31-70) 340-3016

Authorized officer

Cuendet, P



# INTERNATIONAL SEARCH REPORT

International Application No  
PCT/EP2004/051835

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	<p>WO 02/47007 A (EILS ROLAND ; PHASE IT INTELLIGENT SOLUTIONS (DE)) 13 June 2002 (2002-06-13) claims 21,12</p> <p>-----</p>	1-5

# INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/EP2004/051835

Patent document cited in search report		Publication date	Patent family member(s)	Publication date
US 2002077756	A1	20-06-2002	US 6658396 B1	02-12-2003
			US 2004030503 A1	12-02-2004
			US 2003204319 A1	30-10-2003
			US 2003204320 A1	30-10-2003
WO 0237102	A	10-05-2002	AU 1978902 A	15-05-2002
			EP 1328808 A2	23-07-2003
			WO 0237102 A2	10-05-2002
			US 2002155422 A1	24-10-2002
WO 0210453	A	07-02-2002	AU 8088901 A	13-02-2002
			CA 2414421 A1	07-02-2002
			EP 1364049 A2	26-11-2003
			JP 2004522411 T	29-07-2004
			WO 0210453 A2	07-02-2002
			US 2002119462 A1	29-08-2002
WO 0247007	A	13-06-2002	AU 2800002 A	18-06-2002
			CA 2430142 A1	13-06-2002
			WO 0247007 A2	13-06-2002
			EP 1342201 A2	10-09-2003
			JP 2004524604 T	12-08-2004
			US 2004076984 A1	22-04-2004

# INTERNATIONALER RECHERCHENBERICHT

Internationales Aktenzeichen

PCT/EP2004/051835

**A. KLASIFIZIERUNG DES ANMELDUNGSGEGENSTANDES**  
IPK 7 G01N33/50 G06F19/00

Nach der Internationalen Patentklassifikation (IPK) oder nach der nationalen Klassifikation und der IPK

**B. RECHERCHIERTE GEBIETE**

Recherchierter Mindestprüfstoff (Klassifikationssystem und Klassifikationssymbole)

IPK 7 G01N G06F

Recherchierte aber nicht zum Mindestprüfstoff gehörende Veröffentlichungen, soweit diese unter die recherchierten Gebiete fallen

Während der internationalen Recherche konsultierte elektronische Datenbank (Name der Datenbank und evtl. verwendete Suchbegriffe)

EPO-Internal

**C. ALS WESENTLICH ANGESEHENE UNTERLAGEN**

Kategorie*	Bezeichnung der Veröffentlichung, soweit erforderlich unter Angabe der in Betracht kommenden Teile	Betr. Anspruch Nr.
X	US 2002/077756 A1 (DIAMOND CORNELIUS ET AL) 20. Juni 2002 (2002-06-20) Absätze '0050!, '0051!, '0163!, '0217!; Ansprüche 10,21-24	1-5
X	WO 02/37102 A (CHILDRENS MEDICAL CENTER) 10. Mai 2002 (2002-05-10) Absätze '0007!, '0014!, '0015!, '0077!; Ansprüche 11,38-42	1-5
X	WO 02/10453 A (PORTER MARK W ; CASTLE ARTHUR L (US); GENE LOGIC INC (US); JOHNSON KOR) 7. Februar 2002 (2002-02-07) Seite 34, Absatz 1-3 Seite 43, Zeile 28; Ansprüche 51-54,29-43 Zusammenfassung	1-5

-/--

☒ Weitere Veröffentlichungen sind der Fortsetzung von Feld C zu entnehmen

☒ Siehe Anhang Patentfamilie

\* Besondere Kategorien von angegebenen Veröffentlichungen :

\*A' Veröffentlichung, die den allgemeinen Stand der Technik definiert, aber nicht als besonders bedeutsam anzusehen ist

\*E' älteres Dokument, das jedoch erst am oder nach dem internationalen Anmeldedatum veröffentlicht worden ist

\*L' Veröffentlichung, die geeignet ist, einen Prioritätsanspruch zweifelhaft erscheinen zu lassen, oder durch die das Veröffentlichungsdatum einer anderen im Recherchenbericht genannten Veröffentlichung belegt werden soll oder die aus einem anderen besonderen Grund angegeben ist (wie ausgeführt)

\*O' Veröffentlichung, die sich auf eine mündliche Offenbarung, eine Benutzung, eine Ausstellung oder andere Maßnahmen bezieht

\*P' Veröffentlichung, die vor dem internationalen Anmeldedatum, aber nach dem beanspruchten Prioritätsdatum veröffentlicht worden ist

\*T' Spätere Veröffentlichung, die nach dem internationalen Anmeldedatum oder dem Prioritätsdatum veröffentlicht worden ist und mit der Anmeldung nicht kollidiert, sondern nur zum Verständnis des der Erfindung zugrundeliegenden Prinzips oder der ihr zugrundeliegenden Theorie angegeben ist

\*X' Veröffentlichung von besonderer Bedeutung; die beanspruchte Erfindung kann allein aufgrund dieser Veröffentlichung nicht als neu oder auf erfinderischer Tätigkeit beruhend betrachtet werden

\*Y' Veröffentlichung von besonderer Bedeutung; die beanspruchte Erfindung kann nicht als auf erfinderischer Tätigkeit beruhend betrachtet werden, wenn die Veröffentlichung mit einer oder mehreren anderen Veröffentlichungen dieser Kategorie in Verbindung gebracht wird und diese Verbindung für einen Fachmann naheliegend ist

\*Z' Veröffentlichung, die Mitglied derselben Patentfamilie ist

Datum des Abschlusses der internationalen Recherche

12. Januar 2005

Absenddatum des internationalen Recherchenberichts

20/01/2005

Name und Postanschrift der Internationalen Recherchenbehörde  
Europäisches Patentamt, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
Fax: (+31-70) 340-3016

Bevollmächtigter Bediensteter

Cuendet, P

# INTERNATIONALER RECHERCHENBERICHT

Internationales Aktenzeichen  
PCT/EP2004/051835

C.(Fortsetzung) ALS WESENTLICH ANGESEHENE UNTERLAGEN		
Kategorie*	Bezeichnung der Veröffentlichung, soweit erforderlich unter Angabe der in Betracht kommenden Teile	Betr. Anspruch Nr.
A	<p>WO 02/47007 A (EILS ROLAND ; PHASE IT INTELLIGENT SOLUTIONS (DE)) 13. Juni 2002 (2002-06-13) Ansprüche 21,12</p> <p>-----</p>	1-5

# INTERNATIONALER RECHERCHENBERICHT

Angaben zu Veröffentlichungen, die zur selben Patentfamilie gehören

Internationales Aktenzeichen

PCT/EP2004/051835

Im Recherchenbericht angeführtes Patentdokument	Datum der Veröffentlichung	Mitglied(er) der Patentfamilie	Datum der Veröffentlichung
US 2002077756 A1	20-06-2002	US 6658396 B1	02-12-2003
		US 2004030503 A1	12-02-2004
		US 2003204319 A1	30-10-2003
		US 2003204320 A1	30-10-2003
WO 0237102 A	10-05-2002	AU 1978902 A	15-05-2002
		EP 1328808 A2	23-07-2003
		WO 0237102 A2	10-05-2002
		US 2002155422 A1	24-10-2002
WO 0210453 A	07-02-2002	AU 8088901 A	13-02-2002
		CA 2414421 A1	07-02-2002
		EP 1364049 A2	26-11-2003
		JP 2004522411 T	29-07-2004
		WO 0210453 A2	07-02-2002
		US 2002119462 A1	29-08-2002
WO 0247007 A	13-06-2002	AU 2800002 A	18-06-2002
		CA 2430142 A1	13-06-2002
		WO 0247007 A2	13-06-2002
		EP 1342201 A2	10-09-2003
		JP 2004524604 T	12-08-2004
		US 2004076984 A1	22-04-2004